Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 140–144, January 1984
Biophysics

The hydrophobic moment detects periodicity in protein hydrophobicity

(protein structure/α helix/β sheet/3_{10} helix/secondary structure)

DAVID EISENBERG, ROBERT M. WEISS, AND THOMAS C. TERWILLIGER

Molecular Biology Institute and Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90024

Communicated by John T. Edsall, August 1, 1983

ABSTRACT **Periodicities in the polar/apolar character of the amino acid sequence of a protein can be examined by assigning to each residue a numerical hydrophobicity and searching for periodicity in the resulting one-dimensional function. The strength of each periodic component is the quantity that has been termed the hydrophobic moment. When proteins of known three-dimensional structure are examined, it is found that sequences that form α helices tend to have, on average, a strong periodicity in the hydrophobicity of 3.6 residues, the period of the α helix. Similarly, many sequences that form strands of β sheets tend to have a periodicity in their hydrophobicity of about 2.3 residues, the period typical of β structure. Also, the few sequences known to form 3_{10} helices display a periodicity of about 2.5 residues, not far from the period of 3 for an ideal 3_{10} helix. This means that many protein sequences tend to form the periodic structure that maximizes their amphiphilicity. This observation suggests that the periodicity of the hydrophobicity of the protein primary structure is a factor in the formation of secondary structures. Moreover, the observation that many protein sequences tend to form segments of maximum amphiphilicity suggests that segments of secondary structure fold at a hydrophobic surface, probably formed from other parts of the folding protein.**

Ever since the structures of myoglobin and hemoglobin were determined, it has been evident that segments of protein secondary structure can be amphiphilic, in the sense that one side is appreciably more apolar than the other (1, 2). It has also been evident that this amphiphilicity is a factor in the folding of the protein (3).

Recently we have used a measure of the amphiphilicity of a protein segment, the *hydrophobic moment*, to study protein folding (4, 5). Here we extend the application of the hydrophobic moment to the detection of periodicities in the hydrophobicities of amino acid sequences. The specific question we ask is whether known segments of α helix, β structure, and 3_{10} helix tend to be periodic in the hydrophobicities of their amino acid residues. This question can be answered by noting that the hydrophobic moment can give the strength of each component of the periodicity of the hydrophobicity, as shown below.

METHODS

The hydrophobic moment measures the amphiphilicity of the N amino acid residues of a protein segment. If the three-dimensional structure of the protein is known, the hydrophobic moment can be calculated from the relationship

$$\mu_s = \sum_{n=1}^{N} H_n s_n, \qquad [1]$$

in which H_n is the numerical hydrophobicity of the nth residue and s_n is a unit vector in the direction from the nucleus of the α carbon toward the geometric center of the side chain. Protein segments, with most apolar side chains on one side, are characterized by large values of μ_s (5).

The hydrophobic moment of a segment of protein can be estimated when the primary structure is known, provided the segment is periodic and the period is known. Let the periodic structure be specified by m, the number of residues per turn, or alternatively, $\delta = 2\pi/m$, in which δ is the angle in radians at which successive side chains emerge from the backbone, when the periodic segment is viewed down its axis. Thus, for an α helix, δ is 100° ($m = 3.6$), and for a strand of β structure, δ is expected to be in the range of 160° ($m = 2.3$) to 180° ($m = 2.0$). That in the β structure of globular proteins δ might not exactly equal 180° reflects the tendency of β sheets to twist (6, 7). Whatever the value of δ, a periodic structure that is amphiphilic will yield a large value for μ, given by

$$\mu = \left\{\left[\sum_{n=1}^{N} H_n \sin(\delta n)\right]^2 + \left[\sum_{n=1}^{N} H_n \cos(\delta n)\right]^2\right\}^{1/2}, \qquad [2]$$

in which δ is measured in radians—that is, the length of the hydrophobic moment, μ, is given by the sum of the components of the hydrophobicity vectors.

The form of Eq. 2 suggests that we may regard the hydrophobic moment in a more general light: we may think of δ as a variable that may assume any value from zero (for a helix with an infinite repeat) to 180° (for a straight strand of β structure). Then the hydrophobic moment is the modulus of the Fourier transform of the one-dimensional hydrophobicity function,

$$\mu(\delta) = \left\{\left[\sum_{n=1}^{N} H_n \sin(\delta n)\right]^2 + \left[\sum_{n=1}^{N} H_n \cos(\delta n)\right]^2\right\}^{1/2} = \left|\sum_{n=1}^{N} H_n e^{i\delta n}\right|. \qquad [3]$$

Thus, $\mu(\delta)$ is the strength of the component of the periodicity having the frequency δ. This quantity can be evaluated for any amino acid sequence by inserting the associated hydrophobicities in Eq. 3. A large value of μ at a particular δ indicates a strong component of the periodicity at that δ. We note that others have used Fourier transforms to detect residue periodicities in tropomyosin (8) and collagen (9) and that the role of polar and apolar residues in determining protein folding has been discussed by many authors, including Kauzmann (10), Kuntz (11), Rose (12), and Argos *et al.* (13).

Other than for control calculations, the hydrophobicities used in all computations were from the "consensus" scale of ref. 5. These values are as follows: Ile, 0.73; Phe, 0.61; Val, 0.54; Leu, 0.53; Trp, 0.37; Met, 0.26; Ala, 0.25; Gly, 0.16; Cys, 0.04; Tyr, 0.02; Pro, −0.07; Thr, −0.18; Ser, −0.26; His, −0.40; Glu, −0.62; Asn, −0.64; Gln, −0.69; Asp, −0.72; Lys, −1.10; Arg, −1.76.

For examination of periodicities in protein structures, ami-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

no acid sequences (but not atomic coordinates) were taken from the Brookhaven Protein Data Bank, June 1982 tape edition (14). Attention was restricted to those sequences for which authors designated, on the basis of the three-dimensional structure, segments of α structure, β structure, or 3_{10} helix. We attempted to select a representative group of all available structures. Furthermore, the small number of segments of β strands shorter than four residues, of α helices shorter than seven residues, and 3_{10} helices shorter than five residues were ignored. The proteins included in the calculations were as follows:

All-α-helical class: calcium-binding parvalbumin B, cytochrome b_5 (oxidized), cytochrome b_{562} (*Escherichia coli*, oxidized), cytochrome *c* (oxidized), cytochrome c_{551} (oxidized), hemoglobin (horse, aquo met), melittin, cytochrome *c* (prime), myoglobin (met), and myohemerythrin.

α/β class: adenylate kinase, apo-liver alcohol dehydrogenase, carbonic anhydrase form C (carbonate dehydratase), carboxypeptidase A, dihydrofolate reductase, flavodoxin (oxidized form), lactate dehydrogenase (apoenzyme M4), subtilisin novo, triose phosphate isomerase, and L-arabinose binding protein.

All-β class: acid proteinase, endothiapepsin, actinidin (sulfhydryl proteinase), α-chymotrypsin A, concanavalin A, tosyl elastase, papain, Cu–Zn superoxide dismutase, plastocyanin, prealbumin (human plasma), and β-trypsin (diisopropylphosphoryl inhibited).

Additional proteins: phospholipase A2, rhodanese, and thermolysin.

This collection of proteins contains 157 segments of α helix, 220 segments of β structure, and 4 segments of 3_{10} helix meeting the criteria of length described above. These were the subjects of the calculations described in the following.

RESULTS

Model Sequences. To illustrate the application of Eq. 3, the quantity μ was calculated for four model amino acid sequences and plotted as a function of δ (see Fig. 1). This is called a *hydrophobic moment profile*. The first calculation was for a highly amphiphilic α helix, with a sequence arranged such that one side of the helix projects only highly polar arginyl side chains and the other side projects only highly apolar isoleucyl side chains. The profile shows a large maximum at $\delta = 100°$, as well as subsidiary peaks at other angles, which arise because the segment is not infinitely long.

A second calculation was for a model segment of β, Ala-Leu-Ala-Leu-Ala-Leu. This segment was designed such that its maximum amphiphilicity falls at $\delta = 180°$. Furthermore, this maximum is far smaller than that of the α curve because the amphiphilicity of an Ala-Leu sequence is much smaller than that of an Arg-Ile sequence. The Ala-Leu sequence was selected as a model for β because β structure is typically apolar (15).

Two model structures were included in Fig. 1 to give some idea of the factors that cause maxima in the profile at positions other than near 100° or 160–180°, even for α helices and β strands. In an α helix, the maximum can fall away from 100° if the hydrophobic side chains are at positions that form a stripe across the helix at an angle to the axis. Such an arrangement is expected (16, 17) for helices that cross each other at an angle and for which the attractive force is hydrophobic. The curve labeled α cross is for the sequence Ile-Ile-Ala-Ala-Ile-Ile-Ala, which has highly apolar residues at positions 1, 2, 5, and 6 of the α helix. These form a strip of hydrophobic residues ready to pair with another helix having a crossing angle of about 20° with the first (17).

The final model calculation illustrated in Fig. 1 is for a β bulge (18). In a β bulge, the backbone of a strand of β structure is altered so that two successive side chains protrude from the same side of the sheet rather than from opposite sides. The six-residue model for β can be altered to Ala-Leu-Leu-Ala-Leu-Ala to represent a β bulge. In this sequence the strict alternation of a more apolar and a less apolar residue is broken, and the maximum in δ near 180° disappears (Fig. 1, curve labeled bulge).

Not shown in Fig. 1 is the hydrophobic moment profile for a model 3_{10} helix, having a three-residue repeat. This structure forces the main maximum to fall at $\delta = 120°$.

Protein Sequences. When Eq. 3 is applied to the segments of secondary structure listed in *Methods*, the plots of Fig. 2

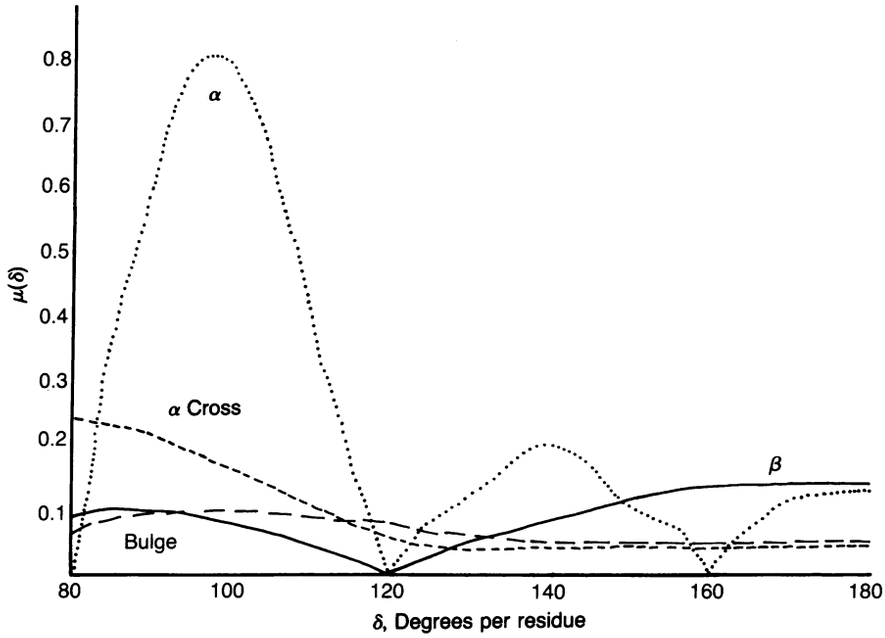

FIG. 1. Hydrophobic moment profiles for model peptides. The values of $\mu(\delta)$ are given here on a per residue basis to allow comparison of peptides. α, Arg-Ile-Ile-Arg-Arg-Ile-Ile-Arg-Arg-Ile-Arg-Arg-Ile-Ile-Arg-Arg-Ile-Ile; α Cross, Ile-Ile-Ala-Ala-Ile-Ile-Ala; β, Ala-Leu-Ala-Leu-Ala-Leu; β Bulge, Ala-Leu-Leu-Ala-Leu-Ala.

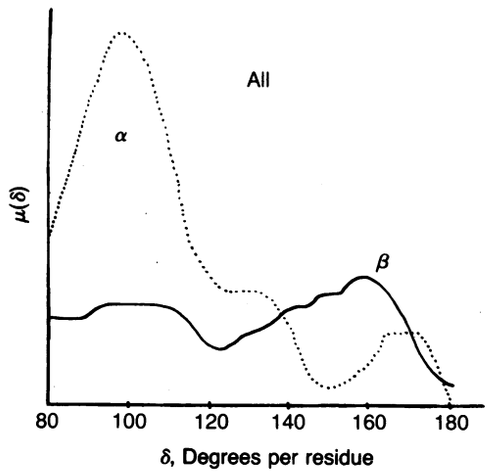

FIG. 2. Hydrophobic moment profiles for protein segments of known structure. For this figure and Fig. 3, the relative ordinates of the different curves are arbitrary. α, The sum of the profiles for 157 segments of α helix. The ratio of the maximum value of $\mu(\delta)$ to the minimum is 1.5. β, The sum of the profiles for 220 segments of β structure. The ratio of the maximum value of $\mu(\delta)$ to the minimum is 1.1.

and 3 are obtained. Fig. 2 shows that the average of 157 segments of α helix results in a strong maximum at 100°. This shows that, on average, these protein primary segments are more polar on one side than on the other when arranged as an α helix (δ = 100°). In fact, on average, these primary structures are more amphiphilic arranged as α helices than in any other periodic secondary structure. For the average of the 220 strands of β structure, the plot shows a maximum at 160°, but it is less pronounced. Part of the reason, as discussed above, is that strands of β structure tend to be more uniformly apolar than α helices.

To explore other reasons why the peak in the β plot is not more pronounced, we divided the proteins considered in the calculation into three main classes of protein structures (19): all α, α/β, and all β (or antiparallel β). The results for the α and β strands in these individual classes are shown in the upper four plots in Fig. 3. From these it is clear that the α maxima are strong in both all-α and α/β type proteins but that the β maxima are more prominent in the β strands of the α/β proteins than in the all-β proteins.

A possible reason for the weak peak at 160° in the all-β proteins is that the β strands in these proteins may be more twisted or more subject to features such as β bulges, which destroy the regularity of the amphiphilicity. This would be expected because the β strands in the all-β proteins tend to be longer (75% are at least seven residues in length, compared to only 42% for the β segments of α/β proteins). To investigate this possibility, we repeated the calculation of the profiles for the β strands of the all-β proteins but truncated each strand after the first six residues. This tends to minimize the effects of features such as β bulges. The results are shown in the lowest curve of Fig. 3 (labeled β6): clearly the maximum is more pronounced than for the corresponding calculation (immediately above it) for the entire lengths of the strands. A similar calculation for the middle six residues of each β strand produced a similar average profile.

Carbonic Anhydrase. When hydrophobic moment profiles are examined for individual segments of secondary structure, many are found to have complex profiles, sometimes with more than a single peak and, in some cases, with local or global maxima far from the expected angle for the given type of secondary structure.

One of the simplest examples is that of carbonic anhydrase, for which the profiles are given in Fig. 4 *Upper* (the α and 3_{10} helices) and *Lower* (the strands of β structure). For none of the four α helices is the maximum exactly at 100°, but three of them (solid curves) fall between 90 and 110°. The longest α helix has its maximum near 130°. The single 3_{10} helix (residues 20–25) has its maximum at 132°, close to the 120° period of an ideal 3_{10} helix. Of the nine strands of β structure, six have their principal maximum between 140 and 180° (solid curves).

Turns, Irregular Segments, and Other Hydrophobicities. As control computations, hydrophobic moment profiles were computed for turns and other irregular segments of polypeptide backbone. Turns were taken as those segments so designated by original authors on the Brookhaven data tape. The sum of the profiles, for the 224 such segments in the proteins listed in *Methods*, is flat relative to the α and β profiles. The same is true for 274 irregular segments, taken from regions of polypeptide chain not designated by authors as any of α, β, 3_{10} helix or turn.

As a final control, the calculations were repeated with three other hydrophobicity scales (5). In all cases the summed hydrophobic moment profiles contained the same essential features.

DISCUSSION

Hydrophobic Periodicity in Proteins. From 25 years of studies of the three-dimensional structures of globular proteins, one of the fundamental observations is that proteins contain periodic structural elements: α helix, β strands, and some 3_{10} helices. Although in these structures, the periodicity of the hydrogen bonding is immediately obvious, Figs. 2 and 3 demonstrate that, on the average, there is also a periodicity in the polar/apolar nature of the amino acid side

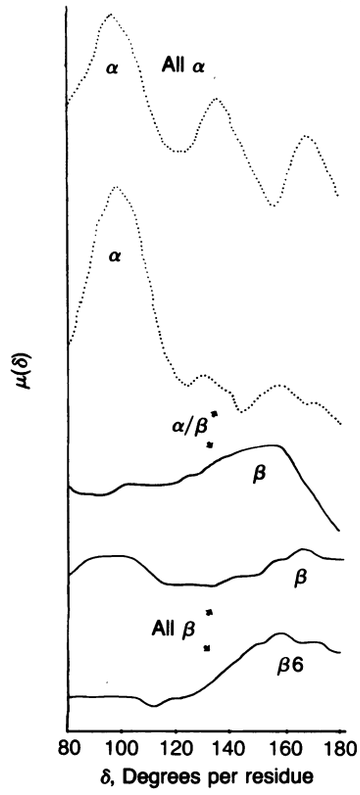

FIG. 3. Hydrophobic moment profiles, similar to those of Fig. 2, except profiles are summed within various structural classes of proteins. The proteins associated with each class are given in *Methods*. The profile labeled β6 is for all-β segments from the all-β proteins truncated to the first six residues.

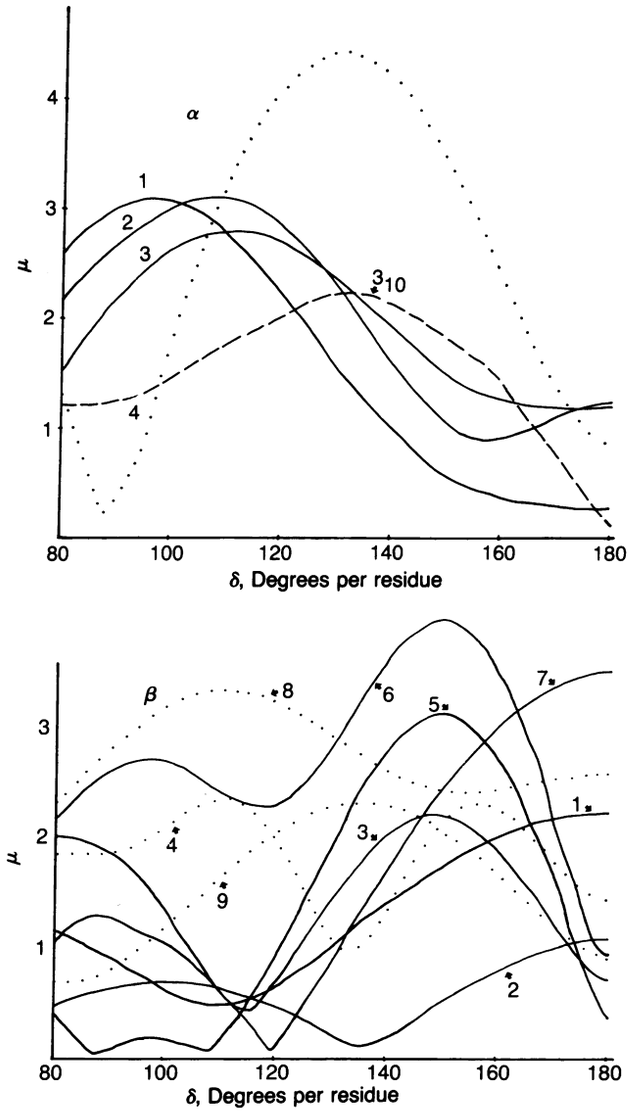

FIG. 4. Hydrophobic moment profiles for secondary structure segments from carbonic anhydrase C. (*Upper*) Helices. Profiles for four α helices and the single 3_{10} helix are shown. The residue numbers of helices are: 1, 130–137; 2, 155–162; 3, 162–168; 4, 219–229; 3_{10}, 20–25. (*Lower*) Profiles for nine strands of β structure. The residue numbers of the strands are: 1, 256–259; 2, 191–196; 3, 206–212; 4, 140–151; 5, 116–124; 6, 87–96; 7, 65–71; 8, 55–62; 9, 171–176.

chains. The periodicity of the hydrophobicity, on the average, matches the periodicity of the structure—that is, α helices on the average tend to have a maximum in the hydrophobic moment profile at $\delta = 100°$, and strands of β structure tend to have a maximum near 160°.

Secondary Structure Formation. Much effort has been focused on the question of what determines the secondary structure of proteins (7, 15). Is it possible that a periodicity in the hydrophobicity encodes a secondary structure with the same period? The results summarized in Figs. 2 and 3 suggest that this could be a factor: a peptide sequence tends to fold into a segment of secondary structure with one apolar and one polar side.

How a peptide sequence might fold into the secondary structure segment yielding the largest possible hydrophobic moment is illustrated by the structure of the melittin tetramer (20, 21). At low concentrations in dilute salt solutions, the 26-residue subunit is a random coil, but at salt concentrations above physiological levels, each subunit of melittin folds into an α helix with an exceptionally large hydrophobic moment (5). In the tetramer, the hydrophobic moments of the four subunits almost exactly cancel. Moreover, helix formation and tetramerization occur at about the same concentration of melittin, as judged by correlated changes in CD and tryptophanyl fluorescence (unpublished observations). Taken together, these observations suggest that the presence of a hydrophobic surface presented by one side of a melittin helix is a favorable energetic environment for the stabilization of other melittin helices. We speculate that this may be a more general aspect of protein folding: protein segments take up secondary structures that yield the largest possible hydrophobic moments because they form at the polar/apolar interface formed between the folding protein and the solution. Nevertheless, it is clear that other factors are also involved in secondary structure formation because by no means do all peptides form the periodic structure having the largest possible hydrophobic moment.

This view of secondary structure formation takes into account "long-range forces," although in a nonspecific way. This is so in that the tendency of a segment to maximize its hydrophobic moment depends not only on the residues of the segment itself but also on those in the rest of the protein sequence. The residues in the other parts must present an apolar surface for the segment to form its proper secondary structure.

Factors Affecting Hydrophobic Periodicity. Hydrophobic moment profiles of segments of protein are available once the primary structure of a protein is known and present the question of what information can be drawn from them. A first step in answering this question is understanding the factors that affect the positions and magnitudes of peaks in the profile.

The positions of peaks in the profile are related in part to the type of secondary structure in the peptide, as demonstrated by the results above. However, not all segments known to be α helices give a profile with a maximum at 100°, nor do all sequences known to fold into β strands give profiles with maxima near 160°. This is evident for helix 4 in Fig. 4 *Upper* and for β strands 4, 8, and 9 in Fig. 4 *Lower* and is even more evident in the corresponding plots for many of the other proteins.

At least two factors that displace maxima from those expected for an amphiphilic structure are discussed above: *(i)* α helices on which the hydrophobic residues are not uniformly distributed along one side parallel to the axis but rather are arranged in a slanted region across the helix and *(ii)* β bulges and other local structures that produce "out-of-phase" periodicities in the distribution of hydrophobic residues. There may well be other reasons that will emerge from detailed studies of the relationship of hydrophobic moment profiles to their corresponding sequences.

The magnitude of the hydrophobic moment profile at a given value of δ is affected by other factors. One is the degree of amphiphilicity of the segment of peptide structure. In general, helices and β segments on the surface of proteins are expected to be quite amphiphilic and hence to have large hydrophobic moments (5, 20, 21). Interior segments are expected to have smaller values.

Longer segments of secondary structure would be expected to have hydrophobic moment profiles revealing fewer strong peaks. The reason is that longer segments are unlikely to have a uniform secondary structure, free of such features as bulges and bends. Thus, the early and late parts of the segment are likely to be out of phase with each other, and, consequently, strong features in the profile tend to be lost.

Another quantity of interest is the value of $\delta(0)$, the value of the hydrophobic moment profile for an infinite period. As is evident from Eq. 3, this is simply the sum of the hydrophobicities of the segment. Thus, the hydrophobicity of a seg-

ment in the present treatment is simply the value of the hydrophobic moment profile at a particular point ($\delta = 0°$).

We acknowledge with thanks suggestions by Drs. R. M. Sweet, D. C. Rees, J. Thornton, and F. R. Salemme and the support of the National Science Foundation, the National Institutes of Health, and a Helen Hay Whitney Fellowship to T.C.T.

1. Perutz, M. F., Kendrew, J. C. & Watson, H. C. (1965) *J. Mol. Biol.* **13,** 669–678.
2. Schiffer, M. & Edmundson, A. B. (1967) *Biophys. J.* **7,** 121–135.
3. Ptitsyn, O. B. & Rashin, A. A. (1975) *Biophys. Chem.* **3,** 1–20.
4. Eisenberg, D., Weiss, R. M. & Terwilliger, T. C. (1982) *Nature (London)* **299,** 371–374.
5. Eisenberg, D., Weiss, R. M., Terwilliger, T. C. & Wilcox, W. (1982) *Faraday Symp. Chem. Soc.* **17,** 109–120.
6. Salemme, F. R. (1981) *J. Mol. Biol.* **146,** 143–156.
7. Schulz, G. E. & Schirmer, R. H. (1979) *Principles of Protein Structure* (Springer, New York), p. 69.
8. McLachlan, A. D. & Stewart, M. (1976) *J. Mol. Biol.* **103,** 271–298.
9. Bear, R. S., Adams, J. B. & Poulton, J. W. (1978) *J. Mol. Biol.* **118,** 123–126.
10. Kauzmann, W. (1959) *Adv. Protein Chem.* **14,** 1–63.
11. Kuntz, I. D. (1972) *J. Am. Chem. Soc.* **94,** 4009–4012.
12. Rose, G. D. (1978) *Nature (London)* **272,** 586–590.
13. Argos, P., Mohana, J. K. & Hargrave, P. A. (1982) *Eur. J. Biochem.* **128,** 565–575.
14. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J. Mol. Biol.* **112,** 535–542.
15. Chou, P. Y. & Fasman, G. D. (1978) *Annu. Rev. Biochem.* **47,** 251–276.
16. Cohen, F. E., Richmond, T. J. & Richards, F. M. (1979) *J. Mol. Biol.* **132,** 275–288.
17. Chothia, C., Levitt, M. & Richardson, D. (1981) *J. Mol. Biol.* **145,** 215–250.
18. Richardson, J. S., Getzoff, E. D. & Richardson, D. C. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 2574–2578.
19. Richardson, J. S. (1981) *Adv. Protein Chem.* **34,** 167–339.
20. Terwilliger, T. C. & Eisenberg, D. (1982) *J. Biol. Chem.* **257,** 6010–6015.
21. Terwilliger, T. C. & Eisenberg, D. (1982) *J. Biol. Chem.* **257,** 6016–6022.